

ABSTRACT

A method for encoding and decoding a sequence is provided. The method comprises searching a set of candidate trees varying in size for a tree T having a plurality of states. Tree T provides a structure that relatively minimizes code length of the sequence from among all the candidate trees. The method further comprises encoding data conditioned on the tree T, which may be a generalized context tree (GCT), using a sequential probability assignment conditioned on the states of the tree T. This encoding may use finite state machine (FSM) closure of the tree. Also provided are methods for decoding an encoded binary string when the encoded string includes a full tree or generalized context tree, as well as decoding an encoded string using incomplete FSM closure, incremental FSM, and suffix tree construction concepts.